Indirect Content Privacy Surveys: Measuring Privacy Without Asking About It

Alex Braunstein*
Chomp
thestatistician@gmail.com

Laura Granka
Google
granka@google.com

Jessica Staddon
Google
staddon@google.com

ABSTRACT

The strong emotional reaction elicited by privacy issues is well documented (e.g., [12, 8]). The emotional aspect of privacy makes it difficult to evaluate privacy concern, and directly asking about a privacy issue may result in an emotional reaction and a biased response. This effect may be partly responsible for the dramatic privacy concern ratings coming from recent surveys, ratings that often seem to be at odds with user behavior. In this paper we propose indirect techniques for measuring content privacy concerns through surveys, thus hopefully diminishing any emotional response. We present a design for indirect surveys and test the design's use as (1) a means to measure relative privacy concerns across content types, (2) a tool for predicting unwillingness to share content (a possible indicator of privacy concern), and (3) a gauge for two underlying dimensions of privacy – content importance and the willingness to share content. Our evaluation consists of 3 surveys, taken by 200 users each, in which privacy is never asked about directly, but privacy warnings are issued with increasing escalation in the instructions and individual question-wording. We demonstrate that this escalation results in *statistically and practically significant differences* in responses to individual questions. In addition, we compare results against a direct privacy survey and show that rankings of privacy concerns are increasingly preserved as privacy language increases in the indirect surveys, thus indicating our mapping of the indirect questions to privacy ratings is accurately reflecting privacy concerns.

General Terms

Human Factors, Design, Security

Keywords

privacy, survey techniques

*This work was done while this author was employed by Google.

Symposium on Usable Privacy and Security (SOUPS) 2011, July 20–22, 2011, Pittsburgh, PA USA

1. INTRODUCTION

Privacy surveys are quite frequent, and the reported results are often dramatic, e.g. more than 70% of users are concerned about online tracking [38] and 93% of users are concerned about company/government access to health records [39]. Almost as frequent though, are reports of consumer behavior that seem incompatible with the high priority on privacy indicated by the surveys. For example, users publicize their purchase histories on web sites like Blippy [4], are willing to trade personal information for entries in sweepstakes or coupons and pay little attention to privacy policies that are reported to be highly valued [16, 15, 18, 35, 9].

We investigate several scenarios in which the mere act of reminding users about general privacy and security issues around content, primes the user. In particular, we study how questions about content use and content importance change when privacy and security language is introduced. As the language escalates, results show increased similarity between responses to our surveys and to a survey that asks about content privacy concerns directly (e.g. "How *private* do you consider this information?").

One explanation for such an effect is education, that is, survey respondents learn, or are reminded, of privacy risk through the survey; thus explaining how behaviors measured elsewhere are inconsistent with survey responses. While some kind of education-effect is likely; it is our belief that it does not account for the bulk of the phenomenon given the high volume of privacy-related news stories in recent years, and the growth of organizations focused on privacy research and privacy breaches (e.g. [32, 10, 28, 26]).

Rather, we argue that explicitly mentioning content sensitivity *invites exaggerated reporting of concerns*. In support of this argument, we discuss three surveys using varying amounts of privacy and security warnings, ranging from an initial survey that does not mention privacy or sensitivity, to one that emphasizes content security/privacy and the risk of content exposure, in the instructions and in some questions. While the survey language includes some security-related issues (e.g. phishing and fraud) the goal to gauge user concern around content exposure, which most users term a privacy concern. In addition, we compare our results with surveys that explicitly ask about privacy. For these reasons, we view our results as most relevant to privacy surveys, although similar effects are likely in the context of security surveys.

The surveys show statistically significant differences in question responses that are consistent with the belief that privacy and security language causes an exaggerated response. For example, the fraction of users willing to share most or all

of their online purchase records with close friends and family decreases by 41% when privacy and security language is introduced in the instructions and questions. In addition, when we use most of the survey results to build models predicting user interest in retrieving a relatively neutral content type, online news, left accidentally in a public place (question 5 in the surveys) we find a marked increase in the retrieving interest as security and privacy language escalates.

Demonstrating the value of indirect surveys for privacy measurement is more difficult, as we are in effect arguing that there is no natural ground truth to compare against. We deal with this challenge by comparing with direct privacy surveys and looking at *rankings* across content types. As has been shown in other domains (see, for example, [23, 33, 2]) rankings are preferable to ratings when measuring values, as they tend to be less sensitive to user-response variations. We show that rankings are similar across surveys, and the survey including the most privacy centered language is most similar to the rankings from a direct privacy survey. This is compatible with the intuition that the increasing amount of privacy language causes the indirect surveys to become increasingly similar to the direct one, and suggests that the rankings from the study with no explicit privacy language may be a more accurate gauge of relative concerns. In addition we show that when mixed models are fit to induce a privacy-oriented ranking on content types, the rankings are more similar to the direct privacy survey results when additional privacy language is present. Finally, we define a "privacy score" which maps responses from our indirect questions to a rating that can be compared with the direct survey results.

Overview. The remainder of the paper is organized as follows. We begin with a discussion of related work (Section 1.1), and then describe our study approach and summarize the content gathered in Section 2. Section 3 demonstrates the impact of escalating privacy language in the surveys on the results with an additional table of analysis in the appendix. We evaluate the usefulness of our approach for measuring privacy concerns in Section 4 and conclude in Section 5.

1.1 Related Work

Many have questioned the accuracy of privacy studies. Organizations like Privacilla.org [30] have raised questions about privacy surveys, some researchers deliberately attempt to obscure their interest in privacy when running studies (e.g. [6]), and several others have noted the discrepancy between reported privacy concerns and user actions (e.g. [15]).

In terms of understanding the impact of privacy survey design, our work most is most closely related to [1] and [21]. In [1], clear inconsistencies between what users report being willing to pay to protect privacy and the compensation they demand in order to give up privacy, are found. In [21], users are primed to give greater credence to privacy through direct manipulations of the survey itself. In particular, half of the survey takers in [21] were provided with a consent warning, hypothesizing that this added emphasis would heighten privacy awareness and lead users to disclose less. Conversely, in another survey variant, the researchers intentionally structured the survey to appear informal, hypothesizing that a casual-looking survey would increase users' comfort in disclosure. We build on both [1] and [21] by providing analogous variations to measure how users report on the importance and use of content types in the presence of escalating priming survey language. Specifically, we hope to better understand relative degrees of priming, by identifying a threshold at which survey language has a measurable effect on responses. In addition, we suggest ways to address the priming effects of survey language by *indirectly* measuring privacy concern.

The challenges of survey wording and analysis are well-studied in other contexts (for example, see, [2, 36, 24]). Our contribution is concrete evidence that these concerns are not only warranted but should certainly exist in the area of content privacy. In particular, we show that providing survey takers with additional privacy and security language will result in responses that are both statistically and practically significant from a baseline.

There are also ongoing efforts to understand how best to design direct privacy studies (see, for example, [22, 7]). In contrast, we present indirect methods for studying content privacy through surveys, and evidence that these methods reduce the exaggerated response possible with direct surveys.

Finally, we note that privacy scores are also used to represent the level of risk undertaken by users of online social networks (e.g. [25, 31]). In contrast, our privacy score aims to represent user privacy concerns around online content.

2. SURVEY DESIGN AND STATISTICAL METHODS

Study	Instructions
1	We are studying the importance of different online information sources in daily life. Please answer a few questions about your use of the given information source or sources.
2	We are studying the importance of different online information sources, many of which are privacy-sensitive and common targets of phishers and others who commit online fraud. Please answer a few questions about your use of the given information source or sources.
3	We are studying the importance of different online information sources, many of which are privacy-sensitive and common targets of phishers and others who commit online fraud. Please answer a few questions about your use of the given information source or sources keeping in mind the potential privacy risks of sharing or otherwise revealing who commit online fraud.

Table 1: Instructions for indirect privacy studies one, two and three.

We conduct two sets of surveys for our analysis, one measuring privacy indirectly and the other directly.

2.1 Indirect Privacy Survey

Our indirect approach to measuring privacy rests on the conjecture that content is sensitive if and only if it is (1) important to the user (aka content owner), (2) important to

Number	Question	Answer Options
1	How frequently do you check [content type]?	Several times a day About once a day A few times a week A few times a month A few times a year Almost never
2	How often do you refer to a [content type] that is several weeks old?	Same as above.
3	How frequently do you forward or otherwise share (e.g. by printing and giving the printed copy) [content type] with your close friends or close family members?	Same as above.
4	(Keeping in mind that purchase records may contain sensitive information,) How many of your [content type] would you be willing to show to your close friends and close family members?	All of them The majority of them Some of them Not very many of them None of them
5	(Keeping in mind that purchase records may contain sensitive information,). if you were to leave a hard copy of one of your [content type] on a restaurant table how likely are you to return to retrieve them?	Very Likely Likely Sometimes I would, sometimes I would not Rarely Never
6	Let's say a server went down and you lost access to your [content type] for two weeks. How would this affect you? It would be...	Extremely disruptive Very disruptive Somewhat disruptive Not very disruptive Not disruptive at all
7	Imagine you have lost access to all the following information sources: email, online calendar, online photos, online documents, Web history and online newspaper. That is, you can no longer access old emails, online calendar entries, online photos, online documents, Web history, online bank/credit card statements, and online newspapers or receive/create new instances of any of these. There is a team available to recover these materials for you, and they need to know how to focus their attention. Please rank the information sources in the order in which the team should work on recovering them (with number 1 being the source the team focuses on first).	Ranked ordering of content types

Table 2: Questions for studies 1, 2 and 3. The parentheticals in questions 4 and 5 were only included in study 3.

Correlation(Q_i,Q_j)	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6
Q_1	1	.66	.51	.09	.25	.69
Q_2	.66	1	.54	.07	.36	.61
Q_3	.51	.54	1	.33	.15	.4
Q_4	.09	.07	.33	1	-.17	-.06
Q_5	.25	.36	.15	-.17	1	.48
Q_6	.69	.61	.4	-.06	.48	1

Table 3: Correlations between question responses in study 1. All correlations are statistically significant (maximum p-value of .006). The correlations between question responses in studies 2 and 3 were quite similar.

at least some others, and (3) infrequently shared. We measure these 3 content dimensions through a series of carefully designed 7 question survey studies. The first 2 studies used identical questions, the last one included additional privacy language in some questions. All the survey questions are in the appendix. The instructions for each survey (see Table 1) include differing levels of privacy related language and warnings, with study one possessing the least and survey three the most. Finally, each study is completed for the following content types: 1) email 2) news 3) online calendar 4) online photos 5) online documents 6) online purchases 7) online bank records 8) web history. Note that our analysis relies exclusively on the ratings-based questions in the studies, questions 1 through 6, as these are easiest to compare (in contrast, question seven requires a ranking, not a rating).

200 users were surveyed for each instruction set/content type combination. There was a small and inconsequential amount of overlap between users for the three groups of instructions. Two users participated in both studies two and three and three users participated in both studies one and two. The users were paid to take part in our study. The users come from a broad pool of testers: the majority of whom have college degrees and are within 24-45 years of age. Only slightly more than half of the pool is male. We do not have demographic information for the specific users who completed our studies. All studies were completed online with no direct interaction between the users and the authors of this paper.

2.1.1 Explanation of Indirect Questions

The survey was designed carefully to address the three necessary characteristics of sensitive content outlined earlier in the section. Questions 1 and 2 assess frequency of use. Questions 3 and 4 are intended to be strong signals of sharing frequency, with the former specifically focused on the importance of content to others – as presumably the user would not share it without cause. Questions 3 and 5 are structured to tap both content importance and content visibility – i.e., how cherished the content is by the user, and to what degree the user is comfortable with others seeing this content. Question 6 speaks most directly to importance of content to the user. Question 7 (summarized in the appendix) serves as a more overt measure of user importance, somewhat replicating the functionality of question 5. The relationships between questions are reflected in Table 3

In Figures 1, 2 and 3 we quickly summarize the aggregate results of questions 4, 5 and 6. We observe an increase in

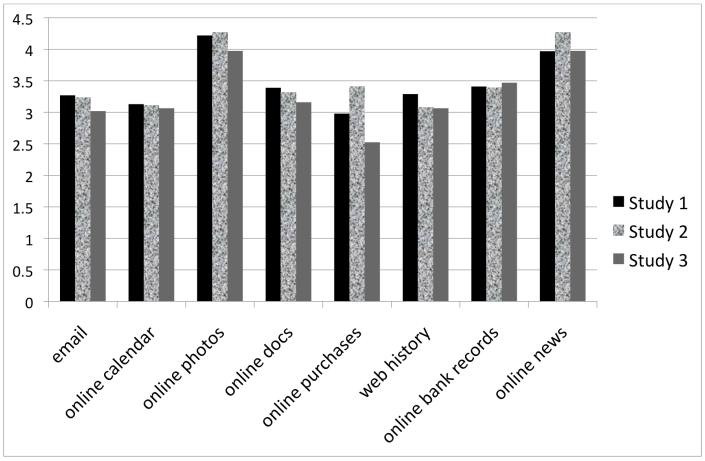

Figure 1: Average Reported Willingness to Share (Question 4) for each content type.

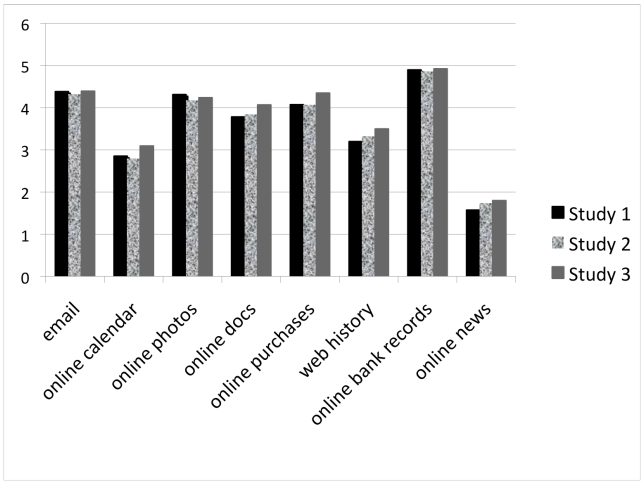

Figure 2: Average Reported Likelihood of Retrieving content Type from Restaurant (Question 5) for each content type.

users reporting they would be likely to retrieve content left at a restaurant across five of the eight content types from weakest (study 1) to strongest (study 3) privacy language in the instructions, a downward trend in reported willingness to share, particularly when moving from study one to study three. Online bank records and email score high consistently on all surveys (users are more likely to retrieve them). Note the stark contrast between online news and the other seven content types. We chose this content type as our control, a non-private baseline to anchor the remaining content types.

The responses to question 6 (Figure 3) do not exhibit a pattern across increasingly strong privacy language in the surveys. This is consistent with our intent that the question measures the importance of the content type to users, rather than an attribute like sharing that has a strong privacy dependency. Similarly, there are no overall trends in frequency of sharing (questions one and two); for a discussion of localized patterns in sharing, see Section 4.2.

Table 2.1.1 shows the rankings for question 7. Note that

Most frequent Choice For Each Ranking	1	2	3	4	5	6	7	8
Study 1	email	bank/CC	docs	docs	purchases	purchases	Web history	news
Study 2	email	bank/CC	docs	purchases	purchases	photos	news	news
Study 3	email	bank/CC	docs	docs	purchases	Web history	calendar	news

Table 4: The most popular content choices for question 7 for each ranked position and each study.

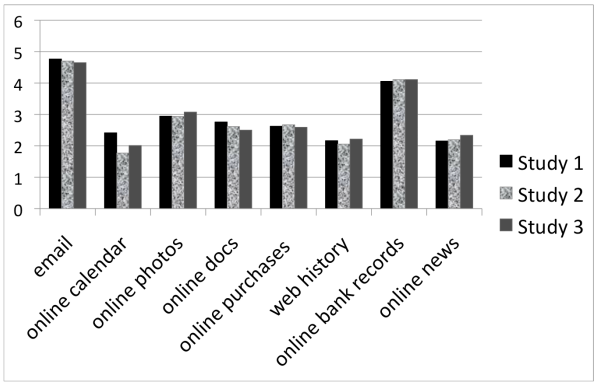

Figure 3: Average Reported Disruptiveness of Lost Access (Question 6) for each content type.

the top 3 choices: email, bank/CC and docs are the same in each study.

2.2 Direct Privacy Survey

In the direct privacy study, users were shown the following instructions:

> Suppose your computer had a virus that gave it access to ALL of your information on this computer and the Internet. Specifically, it would have access to: email, calendar, photos, documents, contacts, Buzz/Twitter, online purchases, web history.

Users were asked four questions about each of the above content types related to: the financial risk of the content, the potential for embarrassment upon content exposure, ease of access to content by the users' contacts and a direct question about privacy: "How *private* do you consider this information?". We use answers to this privacy question, ranked on a five point scale in Section 4.

The survey was taken by 200 users, from the same user pool described in Section 2.1 and overlaps with studies 1, 2 and 3, in the following six content types: email, online calendar, online photos, online documents, online purchases, and web history. The direct survey did not consider bank records or online newspapers. Table 5 summarizes the content privacy rankings relevant to our indirect surveys by average rating.

2.3 Statistical Methods

Our analysis relies on R implementations of traditional tests of significance (i.e. t-tests [14]) and Mixed Effects models to understand the impact of survey wording and to predict question responses. While standard techniques like regression allow for "fixed" effects, such as categorical or indicator variables, their variance estimation is not correct for "random" effects, those that vary within a sampled population hierarchy, hence we utilize Mixed Effects models. Use of these models is quite widespread and dates back to the early 20th century [13]. See [17, 37, 34, 27] for modern and comprehensive treatments.

Before a discussion of our results, it is important to consider the difference between statistical and practical significance in the context of our studies. The former can arise from either large effect *or* large sample sizes, whereas the latter is purely a function of effect size. Defining what effect size is large enough to be practically significant is challenging; we explain this issue in the context of our studies, below.

We most frequently use a 5 point Likert scale. As an example, consider a measured difference of .1 as an effect size. While on its own, this may not seem impressive, if the responses primarily range between 3.25 and 4.5, this difference approaches 10% of the total range and should be considered practically significant. Moreover, several of the responses considered are binary (share or not share, retrieve item with private info from public place, etc.) or represent important semantic differences (e.g. rarely vs. never) and so a small quantitative difference may reflect a meaningful difference in user responses. Consider the logistic curve, a common model of likelihood for binary dependent variables. Different ranges yield very different and quite sizable response variation. While all differences identified as statistically significant are likely not also practically significant, we believe many satisfy this test and encourage our readers to remember these points for context while considering our results.

2.4 Limitations

We discuss the main challenges encountered with our approach to demonstrating the effects of privacy language when measuring privacy concern and discuss how we address them.

First, we rely on self-reporting of behaviors, including prediction of future behaviors. As discussed in more detail in Section 1.1, such self-reporting is notoriously difficult for users. To manage self-reporting errors we built redundancy into our survey. In particular, questions $1, 2, 5, 6$ and 7 all reflect the importance of the content type to the user, question 3 and 4 are both about sharing habits, and questions 3 and 5 reflect the importance of the content to others. All of the questions are used to rank privacy concerns based on these dimensions.

A second important challenge is minimizing response bias that may stem from repeated exposure to the survey questions, either through the completion of multiple studies or as a consequence of answering the survey questions for multiple content types. To minimize such bias we randomized the order of content types within the studies and we ensured that the number of users participating in more than

content Type	Email	Online Documents	Web History	Online Purchases	Online Photos	Online Calendar
Average Privacy Rating	4.32	4.25	3.87	3.76	3.7	2.62

Table 5: Content Privacy Ratings From a Direct Survey, 5 Point Scale

one study was modest (< 2%). In addition, there was at least a 1 week gap between studies.

Finally, we emphasize that we are studying the biasing effect of security and privacy language on survey results. In the presence of such effects we cannot attempt to measure privacy concern absolutely, only relatively. Consider, for example, our neutral category, news, which shows (weak) statistically significant differences in the likelihood of retrieving news content from a public place (question 5) across studies, increasing as the security and privacy language escalates. One possible explanation for the increase is that people genuinely think of news differently in the elevated privacy context – as a reflection of their own political preferences and interests. As news outlets are growing increasingly personalized and arguably polarized, respondents may be sensitive to the notion that others may infer political preferences from news choices.

Another explanation may simply be that the added privacy language in the instructions led users to believe that privacy was ultimately the interesting question for the researchers; this hypothesis might suggest that respondents adjust their responses to accommodate the goals of the experimenter. While either of these explanations may account for the differences in news–the control content type–it is important to note that news remains the type of content that users are least likely to retrieve overall. That is, while interest in retrieving news content increases as security/privacy language is added, its relative ranking amongst other content types remains the same. As previously discussed, the overall privacy ranking of content types may be more reliable than the sheer numerical rating.

3. IMPACT OF SURVEY WORDING

As discussed, the second and third indirect privacy studies include additional privacy-related language in the instructions and question text. We find that these changes to question wording had strong effects, not only on a user's reported willingness to share or retrieve sensitive content, but also with regards to user's self-reported sharing behaviors. The following section provides evidence that introducing additional privacy language impacts user responses about: (1) sharing attitudes, (2) sharing behaviors, and (3) perceptions of data privacy. To better interpret the results, note that the responses listed in Table 2 were mapped onto numerical scores and since there is often a significant difference between the meaning of adjacent responses this can be reflected in an apparently small numerical difference. We try to calibrate the numerical differences in the following.

3.0.1 Known issues in self-report and survey data

It is well known that asking users to self-report their behaviors, particularly about media and data use, can produce unreliable estimates [20, 3, 29]. The lack of accuracy is known to stem from several possible factors: (i) innocent error on behalf of the participant –merely unable to accurately estimate or recall their behaviors, (ii) social-desirability bias – wanting to present the appearance of engage in behaviors that are well-regarded by society, or (iii) priming – intentional or unintentional – due to question wording. As with all surveys, the instrument reported here may unfortunately be a victim of (i), though we do also rely on (iii) to measure significant differences in self-reported behavior survey variants, in the hope to intentionally prime participants with escalating privacy language.

	Study 1	Study 2	Study 3
(Intercept)	4.379	4.316	4.406
	(0.080)	(0.082)	(0.078)
Bank statements	0.518 *	0.533 *	0.536 *
	(0.113)	(0.116)	(0.110)
Calendar	-1.530 *	-1.524 *	-1.297 *
	(0.113)	(0.116)	(0.110)
Documents	-0.601 *	-0.468 *	-0.328 *
	(0.113)	(0.116)	(0.110)
News	-2.808 *	-2.587 *	-2.599 *
	(0.113)	(0.116)	(0.110)
Photos	-0.071	-0.144	-0.170 +
	(0.113)	(0.116)	(0.110)
Purchases	-0.312 *	-0.254 *	-0.041
	(0.113)	(0.116)	(0.110)
Web history	-1.177 *	-0.989 *	-0.893 *
	(0.113)	(0.116)	(0.110)
N	1600	1600	1600
R^2	0.435	0.391	0.414
adj. R^2	0.432	0.388	0.411
Resid. sd	1.134	1.160	1.098

Standard errors in parentheses
* indicates significance at $p < 0.01$

Table 6: Tiered privacy rankings, by study. Linear mixed model results.

3.1 Impact of survey wording on sharing attitudes

We first hypothesize that priming users to think about privacy would encourage them to exhibit more cautious and privacy-conscious responses. Specifically, after being told about the hazards associated with online phishers and fraudsters, users will respond as less willing to share certain types of data.

In fact, we do find this to be the case. Recall the differences in question wording between studies: study 1 asks: "*How many of your [content type] would you be willing to show to your close friends and close family members?*", whereas studies 2 and 3 ask: "*Keeping in mind that [content type] may contain sensitive information, how many of your [content type] would you be willing to show your close friends and close family members?*" Overall, a t-test shows a significant difference between study 1 and 3, with a mean willingness to share content of 3.26 in the first survey, and 3.05 in the third survey (T=4.30, df=3193.513, (p-value < 0.001)).

Studies 1 and 2	Study 3
(1. Bank/CC statements)	(1. Bank/CC statements)
2. Email	2. Email
3. Photos	3. Purchase records
4. Purchase records	4. Photos
5. Documents	5. Documents
6. Web history	6. Web history
7. Calendar	7. Calendar
(8. News)	(8. News)

Table 7: Tiered privacy rankings, by study. We enclose Bank/Credit Card statements, and News in parentheses because these content types were not in the direct privacy survey and so can't be used for comparison purposes.

	Direct Study	Study 3	Study 2	Study 1
1.	email	email	email	email
2.	documents	documents	documents	documents
3.	web history	**purchases**	**purchases**	**purchases**
4.	purchases	**web history**	**photos**	**photos**
5.	photos	photos	**web history**	**calendar**
6.	calendar	calendar	calendar	**web history**

Table 8: Comparison of content Privacy Rankings Between Direct and Indirect Studies. Differences are marked in bold.

	Willingness to Share	Frequency of Sharing
(Intercept)	3.239	1.949
	(0.262)	(0.278)
Study 2	0.087	-0.116 *
	(0.067)	(0.050)
Study 3	-0.172 **	-0.096 *
	(0.056)	(0.045)
N	4800 obs.	4800 obs.
	(626 ids)	(626 ids)
AIC	14286	13477
BIC	14324	13515
LogLikelihood	-7137	-6732
Random Effects Std. Dev		
User	0.701	0.426
Data Type	0.728	0.778
Residual	0.963	0.921

Standard errors in parentheses
* denotes significance at $p < 0.05$; ** $p < 0.01$

Table 9: Sharing behaviors and attitudes: Effects between different survey versions

As described previously, linear mixed effects models are helpful for this between-survey analysis, as they allow us to account for the random variance of certain variables. In this case, we created a linear mixed model with the dependent variable of question 4 – willingness to share, regressed upon the 3-level fixed factor of study (i.e., survey 1, 2, or 3, with survey 1 serving as the baseline). We treated participants as random effects and data types as fixed effects. The model results here indicate no statistical difference between the willingness to share in study 1 and study 2, but in fact a highly significant difference between study 3 and study 1, as comparable with the *t*-tests. Figure 4 shows a partial-effects plot of how willingness to share differs across survey versions. This offers some indication that the privacy wording in study 2 was perhaps not strong enough to produce reactions different from our baseline, though study 3 certainly appears to have emphasized the importance of wording. For full model results, see Table 9.

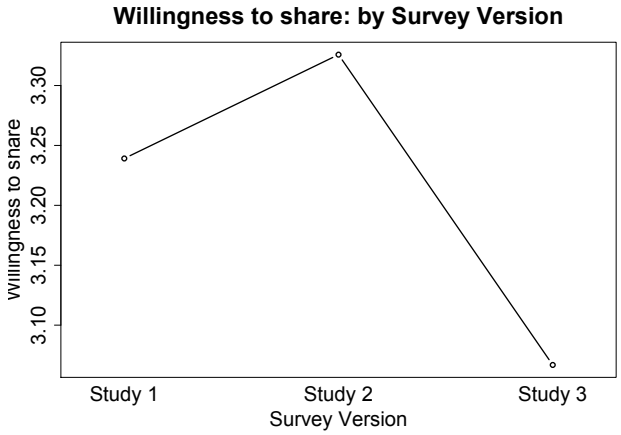

Figure 4: Reported willingness to share between all three survey versions study 1 (neutral), study 2 (privacy warning in instruction), and study 3 (escalated privacy warnings in instruction and question wording). The difference between study 1 and study 2 is within the margin of error.

3.1.1 Privacy language and reactions towards purchase records

Additionally, we are specifically interested in assessing differences between willingness to share online purchase records, as the language used in study 3 may encourage users to think more critically about the personally-identifiable information that purchase records contain (e.g, credit card numbers). In fact, we find that reported willingness to share online purchase records decreased significantly between study 1 and 3: (T = 3.61 (395.32), p-value = 0.0003). The mean willingness to share purchase records was 2.98 in the first survey, and only 2.53 in the third survey. Figure 5 shows user responses for online purchase records according to the 5-level answer options (i.e., 5= "All of them"; 1= "None of them"). As an example of the difference, in the first study 32 users reported a willingness to share all of their online purchase records, whereas in the third study only 10 users had the same willingness. Continuing this trend, while 30 users were not willing to share any online purchase records in the first study, that number grew to 47 in study 3.

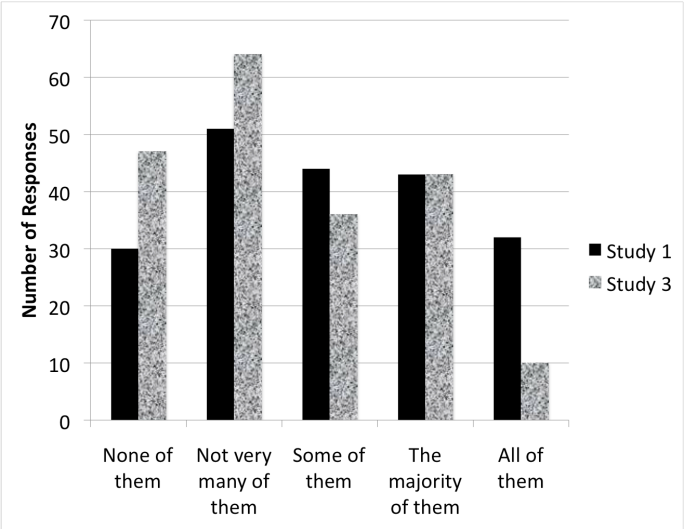

Figure 5: Reported willingness to share online purchase records in study 1 and study 3 (which includes privacy warnings). One a five point scale, the mean value for study 1 is 2.98, whereas for study 3 the mean is 2.525. The difference is significant with a p-value of $< .001$.

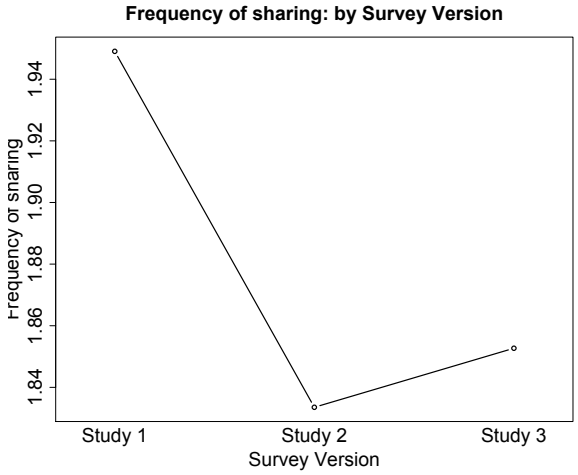

Figure 6: Self-reported frequency of sharing between all three survey versions study 1 (neutral), study 2 (privacy warning in instruction), and study 3 (escalated privacy warnings in instruction and question wording). Both study 2 ($p < 0.05$) and 3 ($p < 0.05$)are significantly different from the baseline study 1.

3.2 Impact of survey wording on self-reported sharing behaviors

In addition to willingness, we also hypothesize that question wording might affect users' self-report sharing behaviors – specifically their reported frequency of sharing content (question 3). As discussed, self-reported behaviors may be influenced by inaccurate memory and recall, and also the social-desirability of being seen to engage in a given behavior. As the survey instructions in studies 2 and 3 explicitly call out the risks of sharing and disclosing data, the additional privacy language may make users believe it socially-desirable to report less of the so-described "risky" sharing behaviors.

Again, we compute linear mixed models, with frequency of sharing (question 3) serving as the dependent variable, and the 3-level factor of study version serving as the independent variable. Participant and data type were included as random effects. The results of this model show that the self-reported frequency of sharing *does* change significantly between studies – with participants in study 2 and 3 reporting significantly lower frequencies of sharing (see Figure 6). The coefficients and other model data are in Table 9.

This result indicates that simply priming people about privacy encourages them to report very different behaviors. What we do not know is whether the privacy language makes participants reflect upon their behavior in a way that aligns with being more privacy-conscious, or whether they report these behaviors in an appeal to social-desirability.

3.3 Impact of survey wording on perceived data privacy

As described, the three studies asked about usage and perceptions of data types, with each study adding an incremental amount of priming about privacy. This section describes how the wording changes uniquely altered participants perceptions of specific data types. This section uses question 5 – which asked users to self-report their likelihood of retrieving forgotten content at a restaurant table – as the dependent measure.

Question 5 was constructed with the aim of capturing more nuanced reactions to privacy, as it encompasses two potential dimensions of privacy – content importance and content sharing. Specifically, retrieving forgotten content is likely to be driven by (1) how important the content is to the user, and (2) how comfortable the user is with others viewing the content. As such, if we see significant differences between the three studies – and in particular between study 2 and 3 – we may be able to better understand the difference between perceived importance and perceived sensitivity when evaluating data privacy.

Overall, we hypothesize that changing the survey instructions in studies 2 and 3 will alter responses to question 5 (retrieval) by increasing user attention to two specific dimensions of privacy: content importance and comfort in content disclosure. As previously discussed, the likelihood of retrieving content will be affected by how much the individual wants the content for *themselves*, and how much the individual desires hiding the content from *others*. Further, we hypothesize that in study 3, with repeated privacy reminders prefacing each question, users will focus more heavily on the dimension of content sensitivity (instead of importance), and as such, reported retrieval rates in study 3 will be higher than in study 2, and certainly in study 1.

3.3.1 Influence of privacy language on control data type: news

All studies included a control content type to determine the baseline effects of question wording on a neutral information source – an online newspaper. As there is no personally-identifiable or sensitive information in a newspaper, one would hypothesize that responses related to this

content type would remain consistent across study iterations. Therefore, if there are significant differences between survey versions for this content type, it is likely attributable to the changes in question wording, and we can again recognize the ability of language to prime users towards more privacy-centric attitudes.

	All Content	Control: News
(Intercept)	2.961	1.281
	(0.076)	(0.144)
Study 2	0.050	0.157 +
	(0.062)	(0.092)
Study 3	0.148 **	0.151 +
	(0.056)	(0.087)
(Q1) Freq. of use	-0.214 ***	-0.031
	(0.016)	(0.033)
(Q2) Freq. of Reference	0.238 ***	0.047
	(0.019)	(0.037)
(Q3) Freq. of Sharing	0.005	0.045
	(0.018)	(0.039)
(Q4) Willingness to Show	-0.216 ***	-0.053 +
	(0.015)	(0.031)
(Q6) How Disruptive	0.520 ***	0.206 ***
	(0.019)	(0.046)
N	4800 obs	600 obs
	626 users	535 users
AIC	15524	1647
BIC	15589	1691
LogLikelihood	-7752	-813.7
Random Effects Std. Dev		
User	0.513	0.637
Residual	1.143	0.677

Standard errors in parentheses
+ denotes significance at $p < 0.1$; ** $p < 0.01$, *** $p < 0.001$

Table 10: Factors affecting content retrieval: Linear mixed models comparing all content types with the baseline of news.

We computed a linear mixed model on the subset of our dataset that is specific to our control – news. Question 5 – rate of retrieving news content – served as the dependent variable, and the independent variables included the 3-level factor of study, along with the remaining survey questions (Questions 1, 2, 3, 4, and 6). Participant was the only random factor. Results for this model show that retrieval rates are comparable in studies 2 and 3, but both are greater than in study 1, with marginal significance (study 2: $t = 1.71$. $p = 0.09$; study 3: t=1.74, $p = 0.08$) (see Figure 8 and view data in Table 10). As news content has minimal to no security risk or personally relevant content, our results suggest that we may confirm our hypothesis that even the minimal wording changes in study 2 were effective – in priming users think more about the *importance* of the content type. That is, we see a priming effect, even in the news category, although to a lower degree than in other content categories.

3.3.2 *Influence of privacy language on all data types*

We again compute a linear mixed model as above, with Question 5 as the dependent variable, but now include all data types in the analysis. Fixed factors again include the 3-level study factor (with study 1 as the baseline), and the survey questions one, two, thee, four, and six as other predictors. Results are displayed in Table 10.

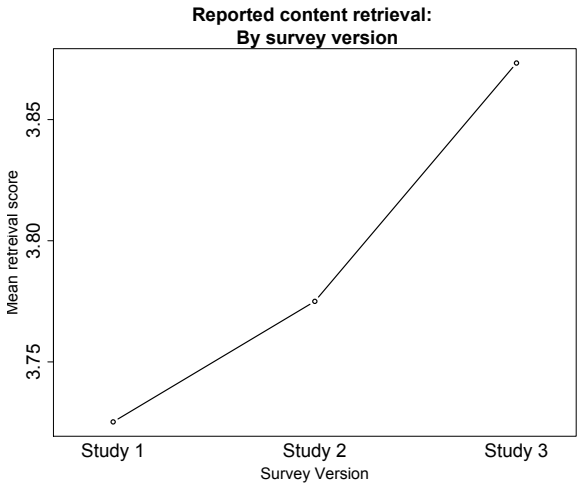

Figure 7: Likelihood of retrieving content, across all three survey versions study 1 (neutral), study 2 (privacy warning in instruction), and study 3 (escalated privacy warnings in instruction and question wording). Study 3 is significantly different from both study 1 ($p < 0.01$) and study 2 ($p < 0.01$).

Overall, results indicate that across all content types, study three, with the most overt privacy language, and hence the strongest priming about privacy concern, showed the highest levels of retrieving content at a restaurant table (see Figure 7). These results again compare with the data presented earlier – that users in study 3 seem to exhibit exaggerated concerns towards privacy.

3.4 Relationship of retrieval to other survey questions

Also interesting to note is the relationship between the remaining survey questions and the likelihood of retrieval. Across all content types, factors that significantly and positively relate to the retrieval include: the frequency of referencing weeks old content and how disruptive a loss of access might be. This is expected, as both of these survey questions tap into how useful and/or critical this content type is. Questions that significantly negatively affected the likelihood of retrieval include the willingness to show others (another expected result, as this effects a user's comfort in having *others* view the content) and the frequency of use (unfortunately, this variable shows effects of model collinearity, which is why we witness a negative coefficient – there is much overlap in what is measured by question 1 and question 2. However, all questions were included in this model to retain an initial comprehensive overall analysis). These relationships are roughly consistent with the (Pearson) correlations between questions in study 1 (Table 3). Note that the self-reported frequency of sharing content had no impact on retrieval.

3.5 Impact of question wording on perceived data sensitivity

We also use question 5 to determine whether privacy word-

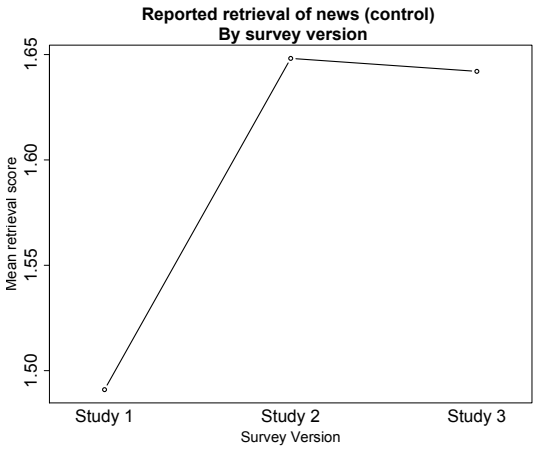

Figure 8: Likelihood of retrieving news content, across all three survey versions study 1 (neutral), study 2 (privacy warning in instruction), and study 3 (escalated privacy warnings in instruction and question wording). Studies 2 and 3 are significantly different (marginally so) from study 1 ($p < 0.08$).

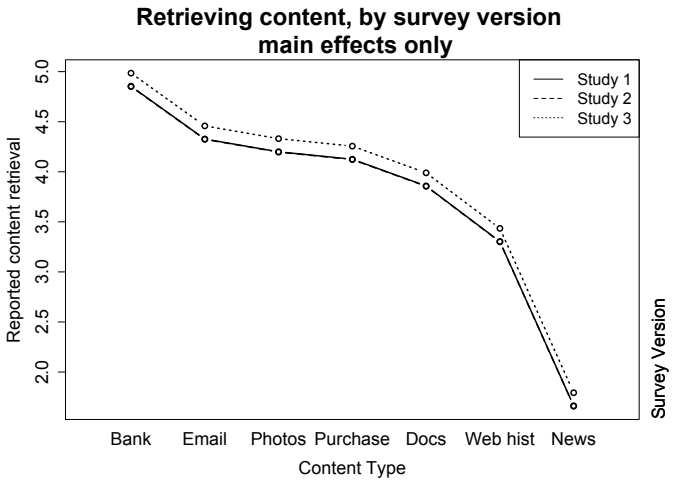

Figure 9: Likelihood of retrieving content, across all three survey versions and by content type. Study 1 (neutral), study 2 (privacy warning in instruction), and study 3 (escalated privacy warnings in instruction and question wording). Study 3 is significantly different from both study 1 ($p < 0.01$) and study 2 ($p < 0.01$).

ing produces significant differences at the level of the individual content type – whether participants respond differently to content types as the privacy wording escalates through the surveys. Specifically, we would expect that in study 3, with repeated privacy reminders prefacing each question, that more sensitive content types – bank statements, purchase records, and documents (where participants reported storing sensitive information) – would be rated more important to retrieve than in study 2, and certainly in study 1. To assess this, we computed models similar to those in Section 3.3, though this time instead of including the different survey questions, we incorporate all content types as fixed factors (with email serving as the baseline content type, as this content was rated most consistently across all study groups). The model we produced predicts the average reported rate of retrieving content at a restaurant table (Question 5), using as predictor variables the 3-level factor of study (study 1 as the baseline), and all eight content types – email, bank statements, calendar, documents, news, photos, purchases, and Web history.

Results from Model 1, which includes only main effects (i.e. independently, how do survey variants and content typed affect the likelihood of retrieval) produce insights similar to those in Figure 7, namely that study 3 produces significantly higher likelihoods of retrieval than the other two studies; see Figure 9). Again, this is strong evidence of the effects of the escalated privacy wording in study 3.

However, the more interesting question we wish to pose is whether our modifications in question wording and degree of priming about privacy only affect participants' reactions towards certain content types. Specifically, while the overt privacy language used in study 3 makes participants *overall* more likely to retrieve content, perhaps these effects predominantly emerge across certain types of content – the content types containing the most personally-identifiable information, such as purchase records and documents (where users reported storing passwords, account numbers, usernames, and other sensitive content).

A further analysis of Question 5 reveals that indeed, the content types which have the potential to pose the highest privacy threat to the user are indeed the ones rated as most likely to retrieve – namely, purchase records, online documents, and Web history (see Figure 10). Bank statements and email remained high privacy concerns across all three surveys, and thus did not significantly differ in effects across survey versions. See full model results in the Appendix 11.

This increased attention to safety or security is also evidenced in the final survey question, where we ask participants to rate which of their lost content types should be restored most immediately from the computer servers. Here again, we see that in the third survey, individuals are very attentive to the content types with personally-identifiable or sensitive information. In study 3, participants were much more likely to attribute bank statements at a higher importance than photos. It is likely that in the first survey – without any privacy indication – we see that individuals respond with more emphasis or thought given to how important or beloved a content type is; e.g., photos are sentimental for people and can re-create memories, etc. However, when participants are primed to think more closely about privacy, their rating is conflated less with the sense of importance or value to the content type, but rather how secure or safe they would feel if the content was disclosed.

4. MEASURING PRIVACY CONCERNS

In Section 3 we argue that privacy ratings are sensitive to survey wording by showing statistically significant differences in average responses as survey wording changes. This highlights the difficulty of evaluating the accuracy of our indirect approach in measuring privacy concern; if privacy ratings are sensitive to wording, then they are not an accurate ground truth to compare against. Instead, we use privacy *rankings* as our ground truth. As in other settings (e.g.

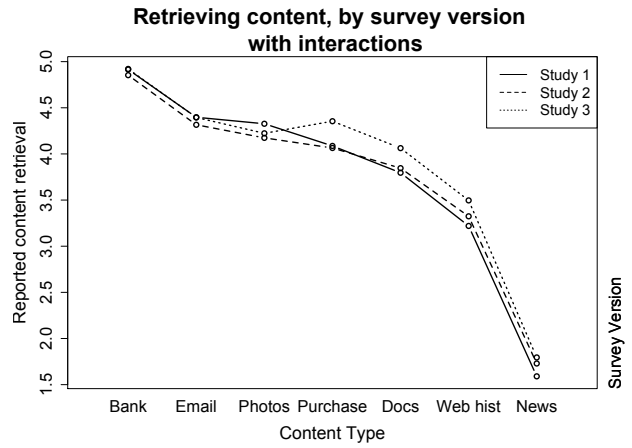

Figure 10: Likelihood of retrieving content, across all three survey versions and by content type. Plot depicts effects of the interaction between survey version and content type. Note that study 3 produces higher retrieval rates for documents, purchase records, and Web history.

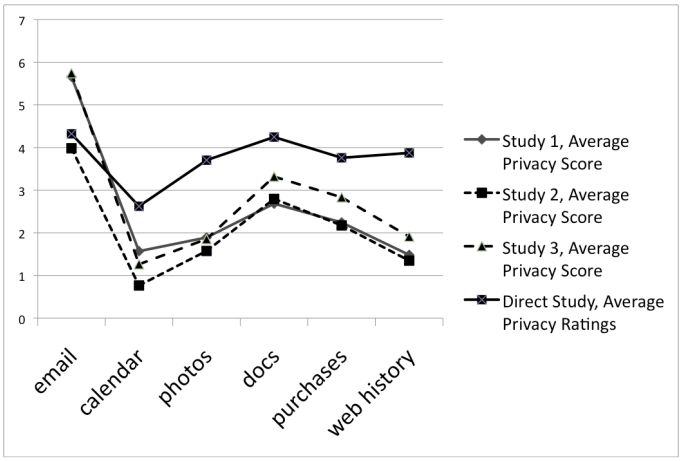

Figure 11: The privacy scores for each of the 3 studies compared against the average privacy ratings reported in the direct study.

[24, 19]) rankings have been found to be more consistent in the face of user variables such as response-style differences and language bias. To achieve this goal, we utilize content rankings from both the indirect and direct privacy surveys described in Section 2 and summarized in Table 5

4.1 Model-Based Rankings

As a first measure of content-specific privacy, we independently model the likelihood of content retrieval (question five) in each of the three studies and compare the resulting content rankings with those from the direct privacy study. We fit a mixed model for each study, predicting likelihood of retrieval, using each content type as a dependent variable, and again using email as the baseline. Note that while question 5 is the particular survey question that is most privacy-oriented, responses to this question may be driven by non-privacy concerns, or simply content desirability/ importance. Further, it is just one survey question, while the entire battery of questions is designed to measure privacy in multiple questions, each addressing one or more dimensions of privacy. Nevertheless, we include this comparison, because even on this single question there are modest differences between the rankings, and it suggests model-based privacy predictions are worth additional exploration. A tiered rank-order for content privacy, based on the beta estimates from these models (in short, we rank in order of beta value), is shown in Tables 6 and 7.

The models reflect some striking differences between survey versions – most notably comparing studies one and 2 with study 3. Recall that as email is the baseline for this regression; the model output presents information on which content types are significantly different from *email*, both positively (more private, e.g., bank statements), or negatively (less private, e.g., news). The content reveals that in studies one and two, the rating of photos is equivalent to the rating for email; however, in study three, photos are rated as less imperative to retrieve (marginally significant; $p = 0.077$). It appears that in study three, with the exaggerated privacy warnings, the importance of photos loses out to the importance of purchase records, which is more compatible with the direct study rankings. These results again support the findings in the prior section – that priming users to think about privacy heightens their attention to the potential privacy risks, and suggests that the absence of such priming language may result in a more accurate gauge.

4.2 Score-Based Rankings

Next, we introduce a ranking mechanism that takes all of our ratings-based survey questions (questions one through six) into account. The advantage of this approach over the model-based one of the previous subsection is that we utilize more survey questions to construct a ranking, and consequently are able to more comprehensively capture privacy concerns. A shortcoming of this approach, however, is that it is more vulnerable to differences in the users in the various study pools. That is, for example, if one pool of users tends to share content a lot more, this may impact our privacy rankings independent of any survey wording changes as scores for one question in one study may be particularly high or low. Regression approaches automatically adjust for this case. In our particular studies, we observed small differences in frequency of checking and sharing online calendars (questions one, two and three)[1]. However, these differences are small and narrowly confined so we do not see any significant impact on rankings derived from these questions.

To compare rankings with our ground truth, we introduce a scoring function based on our original conjecture that sensitive content has three attributes: (1) it is important to the user, (2) it is important to others and (3) it is not readily shared. The scoring function relies on a heuristic for clustering questions into groups using correlation as an indication of association. In particular, the heuristic greedily groups

[1]The difference between studies one and two with respect to online calendars is modest (means of 1.51 and 1.23 respectively on question three, with p-value of .002). Users in study one also check their online calendars (question one, means of 2.76 and 2.155, respectively, with a p-value of .0006) and refer to old calendar entries (question two) more frequently than those in study two (question two, means of 1.785 and 1.545, respectively, p-value of .02).

questions into potentially overlapping clusters according to their correlation by forming, for each question, the largest cluster of questions such that all question pairs, Q_i, Q_j in the cluster have correlation (Pearson correlation, for example [14]) of at least some threshold value r.[2] Note that this rule may induce overlapping clusters, which are compatible with the intuition that questions may reflect multiple attributes of privacy.

To make this heuristic concrete, we measure the correlations between distinct question pairs (Table 3) and cluster with a threshold equal to the median correlation, .4. Using this heuristic we extract the following overlapping clusters: $\{Q_1, Q_2, Q_3, Q_6\}$, $\{Q_4\}$, and $\{Q_5, Q_6\}$. Note that the second cluster, $\{Q_4\}$ represents the sharing attribute as it corresponds to the question, "...how many of your [content type] would you be willing to show to your close friends and close family members?". The first and last clusters both seem to represent importance of the content type, with the first cluster perhaps more weighted toward the importance of the content type to the user, and the third cluster more evenly balanced between importance to the user and importance to others.

We use these clusters to build the following privacy score, in which A_i denotes the average response to question Q_i on a 0-5 scale (normalized from a six point scale in the case of questions 1, 2, and 3).

$$P(\{A_i\}_{i=1,\ldots,6}) = \frac{1}{4}(A_1 + A_2 + A_3 + A_6) - A_4 + \frac{1}{2}(A_5 + A_6)$$

The privacy scores from each study are shown in Figure 4 with the privacy ratings from the direct privacy study 5.

We expect the rankings from our studies to be monotonically increasing in similarity (thought not strictly) to the direct privacy study. Table 8 reflects this. Relative to the direct privacy ranking, study 3 makes 2 errors, study 2 makes 3, and study 1 makes 4.

We also observe this similarity in Pearson correlation ranking between each study and the direct privacy study. Though the correlation between the average privacy scores in study one and the direct privacy survey is not statistically significant (correlation of .59, with p-value of .21), once privacy language is added, the correlations grow and are at least weakly significant. Specifically, the correlation between study two and the direct study is .813, with p-value of .049 and between study three and the direct study the correlation is .73 with a p-value of .097.

5. CONCLUSION AND OPEN PROBLEMS

We have presented statistically significant evidence that privacy survey wording strongly impacts responses by increasing user reports of privacy concern both with respect to relatively innocuous content types (e.g. news articles) as well as content that contains personal information (e.g. purchase records). We also suggest mechanisms for translating responses to indirect questions into privacy ratings and show that this mapping increasingly preserves relative rankings of content types from direct privacy surveys, as more privacy language is introduced.

We've taken first steps toward a methodology for indirect privacy surveys; much work remains. We highlight three problem areas:

1. Comparing survey results with user behavior: We've shown that with escalating language reported results between indirect and direct surveys become increasingly similar, but what can we say about the indirect survey results with little or no privacy language? For example, how consistent are the results of study 1 with current user practices?

2. Privacy scores: We've identified a basic and somewhat ad hoc heuristic. More principled statistical approaches a la Principal Components Analysis (PCA), factor analysis, or other matrix decomposition and linear projection methods may prove fruitful. These were explored in our data analysis, but a discussion is beyond the scope of this manuscript.

3. Algorithmic support for generating indirect surveys: We've focused on content privacy. For other privacy problems, is there a principled way to go from the problem to a set of associated attributes (like importance and sharing in the case of content privacy)?

6. ACKNOWLEDGMENTS

The authors are very grateful to Daniel Abizeid for help with the studies.

7. REFERENCES

[1] A. Acquisti, L. John and G. Lowenstein. What is privacy worth? Workshop on Information Systems and Economics (WISE), 2009.

[2] D. Alwin, J. Krosnick. The Measurement of Values in Surveys: A Comparison of Ratings and Rankings. Public Opinion Quarterly, Vol. 49, No. 4 (Winter, 1985), 535-552.

[3] J. Barabas and J. Jerit. Are Survey Experiments Externally Valid? American Political Science Review 104 (May): 226-42. 2010.

[4] Blippy. http://blippy.com/

[5] T. Buchanan, C. Paine, A. Joinson and U-D. Reips. Development of Measures of Online Privacy Concern and Protection for Use on the Internet. Journal of the American Society for Information Science and Technology, 58(2):157-165, 2007.

[6] L. Brandimarte, A. Acquisti and G. Lowenstein. Misplaced Confidences: Privacy and the Control Paradox. WEIS 2010.

[7] T. Buchanan, A. Paine, A. Joinson and U. Reips. Development of Meausures of Online Privacy Concern and Protection for Use on the Internet. Journal of the American Society for Information Science and Technology, 58(2): 157-165, 2007.

[8] [etc] A frustrated user lashes out at Google after being burned by the launch of Buzz and attendant loss of privacy. ArsTechnica, February 12, 2010.

[9] M. J. Culnan and G. R. Milne. The culnan-milne survey on consumers and online privacy notices, 2001. http://intra.som.umass.edu/georgemilne/pdf files/culnan-milne.pdf.

[10] Electronic Frontier Foundation. http://www.eff.org/

[2]Question seven, a ranking question, is not used by the heuristic because it is fundamentally different and less straightforward to transform to a numeric scale.

[11] S. Egelman, J. King, R. Miller, N. Ragouzis and E. Shehan. Security User Studies: Methodologies and Best Practices. CHI 2007.

[12] Facebook + Husband = grrrrrrrr!. Anonymous post at iVillage. http://forums.ivillage.com/t5/Girl-Talk/Facebook-Husband-grrrrrrrr/td-p/116924103 February 23, 2011.

[13] R. A. Fisher. The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh 52: 399433. (1918).

[14] D. Freedman, R. Pisani, R.Purves. Statistics. W. W. Norton and Company, 1997.

[15] J. Gideon, S. Egelman, L. Cranor and A. Acquisti. Power Strips, Prophylactics and Privacy, Oh My! SOUPS 2006.

[16] E. Goldman. The Internet Privacy Fallacy. http://eric_goldman.tripod.com/articles/privacyfallacy.htm

[17] H. Goldstein. Multilevel Statistical Models, Second Edition, London: Edward Arnold. 1995.

[18] Harris Interactive. Identity Theft: New Survey & Trend Report, August 2003.

[19] A-W. Harzing. Rating Versus Ranking: What is the Best Way to Reduce Response and Language Bias in Cross-National Research. International Business Review, Volume 18, Number 4, 2009.

[20] C. L. Hovland. Reconciling Conflicting Results Derived From Experimental and Survey Studies of Attitude Change. American Psychologist, 14: 8-17. 1959.

[21] L. John, A. Acquisti and G. Lowenstein. Strangers on a place: Context-dependent willingness to divulge personal information. Journal of Consumer Research, 2010.

[22] A. Joinson, C. Paine, T. Buchanan, and U-D. Reips. Measuring self-disclosure online: Blurring and non-response to sensitive items in web-based surveys. Computers in Human Behavior, Volume 24, Issue 5, September 2008, pp. 2158-2171.

[23] M. L. Kohn. Reassessment 1977. In Kohn, Class and Conformity: A Study in Values. 2nd ed. Chicago: University of Chicago Press.

[24] J. Krosnick and D. Alwin. A Test of the Form-Resistant Correlation Hypothesis: Ratings, Rankings and the Measurement of Values. Public Opinion Quarterly, Vol. 52, No. 4 (Winter, 1988), pp. 526-538.

[25] K. Liu and E. Terzi. A framework for computing the privacy scores of users of online social networks. ICDM 2009.

[26] Pew Research Center. http://pewresearch.org/

[27] J. C. Pinheiro and D. M. Bates Springer. Mixed-Effects Models in S and S-PLUS. ISBN 0-387-98957-9, 2000.

[28] Ponemon Institute. http://www.ponemon.org/news-2/40

[29] M. Prior. The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. Public Opinion Quarterly, 73 (1): 130-143. 2009.

[30] Privacilla.org. Privacy Survey Design is Often Flawed. Available at: http://www.privacilla.org/fundamentals/surveyqs.html

[31] Privacy Check. http://www.rabidgremlin.com/fbprivacy/

[32] Privacy Rights. http://www.privacyrights.org/

[33] M. Rokeach. The Nature of Human Values. New York: Free Press, 1973.

[34] S. R. Searle, R. Casella, and C. E. McCulloch. Variance Components. New York: John Wiley and Sons. 1992.

[35] S. Spiekermann, J. Grossklags, and B. Berendt.E-Privacy in 2nd Generation E-Commerce: Privacy Preferences versus Actual Behavior. In Proceedings of EC01: Third ACM Conference on Electronic Commerce, pages 3847, Tampa, Florida, 2001. http://www.sims.berkeley.edu/ jensg/research/ eprivacy acm.html.

[36] E. Shaeffer, J. Krosnick, G. Langer and D. Merkle. Comparing the Quality of content Obtained by Minimally Balanced and Fully Balanced Attitude Questions. Public Opinion Quarterly, Vol. 69, No. 3, Fall 2005, pp.417-428.

[37] T. A. B. Snijders and R. J. Bosker. Multilevel Analysis An Introduction to Basic and Advanced Multilevel Modeling. London: Sage. 1999.

[38] J. Turow. Americans & Online Privacy: The System is Broken. A report from the Annenberg Public Policy Center of the University of Pennsylvania. June 2003.

[39] United Press International/Zogby. UPI Poll: Concern on health privacy. February 21, 2007. http://patientprivacyrights.org/2007/02/upi-poll-concern-on-health-privacy/

APPENDIX

	Main Effects	Interactions
(Intercept)	4.334	4.405
	(0.058)	(0.078)
Study 2	-0.009	-0.081
	(0.061)	(0.109)
Study 3	0.139 **	-0.006
	(0.053)	(0.105)
Banking	0.527 **	0.517 **
	(0.056)	(0.097)
Calendar	-1.452 **	-1.532 **
	(0.056)	(0.097)
Documents	-0.469 **	-0.602 **
	(0.056)	(0.097)
News	-2.667 **	-2.808 **
	(0.056) **	(0.097)
Photos	-0.129 **	-0.071
	(0.056)	(0.097)
Purchases	0.204 **	-0.312 **
	(0.056)	(0.097)
Web history	-1.021 **	-1.177 **
	(0.056)	(0.097)
Study 2 * Banking	—	0.018
	—	(0.137)
Study 2 * Calendar	—	0.009
	—	(0.137)
Study 2 * Documents	—	0.136
	—	(0.137)
Study 2 * News	—	0.223
	—	(0.137)
Study 2 * Photos	—	-0.070
	—	(0.137)
Study 2 * Purchases	—	0.061
	—	(0.137)
Study 2 * Web history	—	0.189
	—	(0.137)
Study 3 * Banking	—	0.011
	—	(0.138)
Study 3 * Calendar	—	0.234 +
	—	(0.138)
Study 3 * Documents	—	0.266 *
	—	(0.138)
Study 3 * News	—	0.204
	—	(0.138)
Study 3 * Photos	—	-0.103
	—	(0.138)
Study 3 * Purchases	—	0.266 *
	—	(0.138)
Study 3 * Web history	—	0.280 *
	—	(0.138)
N	4800 obs.	4800 obs.
	(626 ids)	(626 ids)
AIC	14164	14207
BIC	14241	14376
$LogLikelihood$	-7070	-7078
Random Effects Std. Dev		
User	0.587	0.587
Residual	0.968	0.967

Standard errors in parentheses
$^{+}$ denotes significance at $p < 0.1$; ** $p < 0.001$

Table 11: Likelihood of Retrieval: Effects of question wording and content type